Face Alignment Using Local Probabilistic Features

Qing Lu[1], Jun Yu[1], and Zengfu Wang[1,2(✉)]

[1] Department of Automation, University of Science and Technology of China, Hefei, China
ustclq@mail.ustc.edu.cn, {harryjun,zfwang}@ustc.edu.cn
[2] Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China

Abstract. Random forest is an effective tool for locating facial landmarks. In this paper, we propose a novel random forest based face alignment method using local probabilistic features (LPF). Here, the LPF has the property of calculating the probability of a sample belonging to the leaf nodes of a tree. The obtained LPF is then used to train a regression model for approximating to the real location of facial landmarks. The above procedure is repeated several times step-by-step in a cascaded form until the model converges. In the end, various convergent results are combined to overcome the instability of a single one. By this way, our method markedly outperforms the state-of-the-art methods. Experimental results show the effectiveness of our algorithm on various face alignment datasets.

Keywords: Random forest · Face alignment
Local probabilistic features · Cascaded form

1 Introduction

Face alignment aims to locate facial landmarks, such as mouth corners, nose tip, pupil and chin. It is a fundamental component in many applications (e.g., facial attribute inference [15], face recognition [13], face verification [17] and facial animation [18]). With the rapid growth in image data nowadays, a highly efficient and accurate face alignment method is in great demand. Though great success has been achieved in this field, accurate and robust alignment of facial landmarks is still a formidable challenge due to partial occlusion and large variations of head pose.

Active appearance model (AAM) [7] solves the task of face landmarks detection by reconstructing entire face using an appearance model and minimizing model parameter errors in training phase. The model of AAM may not work well on unseen faces as the limited expressive power to model texture space. It is also well known that AAM can not handle large variations of expression, illumination and initialization. To solve this problem, local feature based methods such as active shape model (ASM) [8] and constrained local model (CLM) [9]

B. Zeng et al. (Eds.): PCM 2017, LNCS 10735, pp. 184–193, 2018.
https://doi.org/10.1007/978-3-319-77380-3_18

have been proposed, which only model the local appearance around the landmarks instead of the entire face. The results show better generalization ability and stability performance. However, the local features sampled from the current facial landmarks are still not robust enough to adapt to large deformation, pose variation and occlusions.

The vast majority of face alignment approaches proposed in recent years are on the basis of shape regression [6,10,11,14,21]. The advantages of these methods are reflected in the ability of adaptively enforcing shape constrains and the capability of effectively leveraging large bodies of training data. The shape regression algorithm is frequently used in a cascaded manner. Cascaded shape regression (CSR) is first put forward in [6]. Without using a fixed parametric shape model, the inherent shape constraint in [6] is encoded into a cascaded regression framework and implemented from coarse to fine during the test phase. Beginning with an initial shape calculated from the average facial landmarks of the training datasets, the face shape is optimized stage-by-stage by adding a shape increment. In each stage, features are extracted from the images and then used in a regression method to calculate the current location of the facial landmarks.

The selection of features is crucial to the results of regression, so a series of algorithms on it is gradually put forward. The efficiency can be obviously improved by employing the shape-indexed feature [6]. In [26], SIFT (scale-invariant feature transform) feature is used to achieve a robust representation against illumination. Sun et al. [24] takes the advantage of the deep structures of convolutional networks to learn the features. In [21], local binary features (LBF) are presented for extremely accurate and fast face alignment. The obtained LBF is incorporated into CSR framework to learn a linear regression. Due to the simplicity of the pixel based feature, LBF is an exceedingly efficient tool for facial landmarks location. Nevertheless, it is more sensitive to noise compared with other conventional methods, such as HOG (histogram of oriented gradient) and SIFT.

In this paper, we propose a method using local probabilistic features (LPF), which is an optimization of LBF. The proposed LPF has the ability of modeling the probability of a test sample belonging to each leaf node. In the process of tree node split, we employ the average pixel difference value of three pairs of pixels, which can not only guarantee the accuracy of the algorithm, but also improve the speed of the algorithm. In order to obtain the optimized output results, various convergent models are combined to contribute their respective advantages.

The main contributions of our method are:

1. As an extension of local binary features [21], we focus on the important role of the probability for improving the learning effectiveness and efficiency in random forest at the first time. The method synthesizes not only the efficient performance of LBF, but also the probability of a sample reaching each leaf node. Qualitative and quantitative results show the superiority of our algorithm by blending them together.

2. Traditionally, the results of facial landmarks detection are determined by a single regression model [6,10,14,21]. We overcome this limitation by combining various convergent models, since each model has its unique advantages. By integrating them together, we can overcome the instability of a single one.

2 Related Work

2.1 The Cascade Shape Regressors

In recent years, the concept of cascade shape regression gradually shows its superior quality in the research field of face alignment. All these methods take face alignment as a regression problem. Cascade shape regressors generally employ N regressors in series form. The vector S consists of the x, y-coordinates of L facial landmarks. Beginning with an image and a raw initial face shape S^0, S is optimized by a shape increment δS^n, which is calculated by the regressor R^n, $\mathrm{n} = 1, 2, ..., \mathrm{N}$, stage-by-stage:

$$S^n = S^{n-1} + \delta S^n \tag{1}$$

where S^n is the current shape estimation, S^{n-1} is the shape estimated by the previous stage, δS^n is calculated as follow:

$$\delta S^n = W^n \Phi^n(I, S^{n-1}) \tag{2}$$

where W^n is a matrix for global linear projection, Φ^n is a feature mapping function.

2.2 Random Forest

In recent years, random forests [4] play a great role in many classic pattern recognition problems, such as image classification [3], data clustering [23] and shape regression [21]. This approach has many advantages: (a) efficiency in both training and prediction, (b) the ability to handle a large number of input variables, (c) the ability to detect the interaction between features, and (d) suitable for multi-classification problem.

3 Method

In the training phase, we first augment our training data to meet the diversity of different situations. Then the local probabilistic features are generated from the random forest. After that we learn a global linear projection W^n by dual coordinate descent method [12]. The above procedure is repeated N times step-by-step in a cascaded form. The overview of our training algorithm is shown in Table 1.

The proposed local probabilistic features are extracted from conventional random forest. Following the framework proposed by [20], one facial landmark

Table 1. Training of cascade face alignment with local probabilistic features

Algorithm 1 Training of cascade face alignment with local probabilistic features

Input: training datasets (images I_i, ground truth shapes $\widehat{S}_i$ and initial shapes S_i^0), the number of stages N, the numbe of trees in each stage T, the tree depth D, the number of landmarks L
Output: W^n, Φ^n
for n=1 to N **do**
1: expand training samples
augment training samples with multiple initializations for one image, use an augmentation of 5 times
2: train random forest
follow the framework proposed by [4]
3: extract local probabilistic features
for l=1 to L **do**
extract local probabilistic features ϕ_l^n
end for
Φ^n=[ϕ_1^n; ϕ_2^n; ...; ϕ_L^n]
4: learn the global regressionW^n
follow the framework proposed by [21]
5: update shapes using W^n, Φ^n of this stage
$\boldsymbol{S}_i^n$=$\boldsymbol{S}_i^{n-1}$+$W^n\Phi^n(I,S^{n-1})$
end for

corresponds to one random forest, which is composed of 10 trees. For each leaf node, we calculate the local probabilistic features as follow:

$$p'(i) = \frac{num(i)}{\sum_{i=1}^{I} num(i)} \tag{3}$$

where $p'(i)$ is the initial probability value of the *ith* leaf node of a tree, $num(i)$ is the number of training samples falling into the leaf node.

$$p(i) = \begin{cases} lowTh, & \text{if } p'(i) < \text{lowTh} \\ p'(i), & \text{if lowTh} \le p'(i) < \text{highTh} \\ highTh, & \text{if } p'(i) \ge \text{highTh} \end{cases} \tag{4}$$

where $p(i)$ is the final probability value of the *ith* leaf node of a tree, *lowTh* and *highTh* are the lower threshold and the upper threshold, respectively.

For each facial landmark, we train a random forest through the method proposed by [4]. Figure 1 roughly illustrates the process of producing local probabilistic features from random forest. Each leaf node contains a pixel difference feature f [6], a local probabilistic feature $p(i)$, and a threshold.

When all of the local feature mapping functions ϕ_i^n, $i = 1, ..., \mathrm{L}$, have been established, high-dimensional probabilistic features are formed by concatenating all local probabilistic features. Details are shown in Fig. 2.

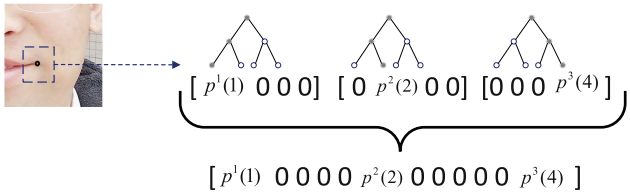

Fig. 1. The process of producing local probabilistic features from random forest. For the parameter $p^i(j)$, i denotes the ith tree in its random forest and j denotes the jth leaf node in its tree.

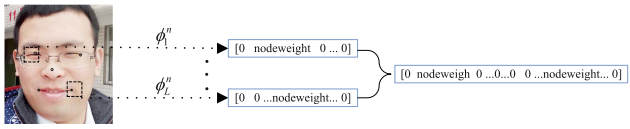

Fig. 2. The high-dimensional probabilistic features are formed by concatenating all local probabilistic features.

The process of testing is same as that for training, but we add an optimization strategy by combining various models to overcome the instability of a single one.

$$S = \frac{1}{M}\sum_{i=1}^{M} S_i = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{L} s_i(x_j, y_j) \tag{5}$$

where M is the number of models, S_i is the shape calculated by the ith model, $s_i(x_j, y_j)$ denotes the location of jth landmark.

4 Experiments of Alignment

In this section, we first confirm the selection of some parameters, which are critical to our experimental performance. Then, we show the effectiveness of our method on two datasets, 300 W and Helen.

4.1 Datasets

Helen [16] consists of 2000 training and 330 test web images. The high resolution images are useful for accurate location. In order to achieve rapid results, we employ 68 landmarks instead of 194 landmarks to show the performance of our method. **300 W** [22] is short for 300 faces in the wild. It is created from classical datasets, including LFPW [2], Helen [16], AFW [20], XM2VTS [19] and IBUG. Our training images are composed of the training sets of Helen and LFPW, and the whole AFW dataset. Our testing images are composed of the test sets of LFPW and Helen, which are also called the common test set, and the whole IBUG dataset, which is also called the challenging test set.

4.2 Selection of Parameters

In our experiments, a multi-pose V.J. detector is used for detecting face rectangles, and the mean shape of the training data is chosen as the initial shape of test image.

For the proposed method, the number of cascade stages N and the number of trees T in every random forest are crucial parameters. Figure 3 shows the mean errors as a function of the number of cascade stage. We can see that when N increases to 7, the performance of the algorithm achieve an ideal state. Figure 4 shows the mean errors as a function of the number of trees. With the growth of this parameter, the performance is fluctuating and the test time is increasing.

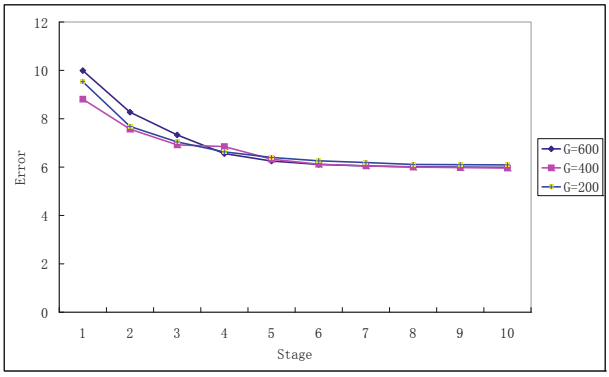

Fig. 3. The mean error at each stage of the cascade is plotted. Using many stages of regressors is fairly useful, regardless of the number of dot group, G, in each forest.

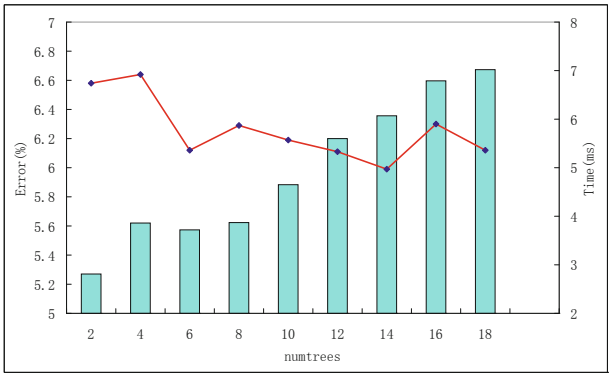

Fig. 4. The mean error as a function of the number of trees T in each random forest (line chart). The test time as a function of T (bar chart).

4.3 Results and Discussion

Our method is implemented in C++ and tested on a i7-6700 CPU. The results of other compared approaches are from the reports in original papers. When the parameter T is set to 2, we can see that the alignment rate of our method is better than others from Table 2.

To evaluate the effectiveness of the local probabilistic features, we take our experiments on two frequently-used datasets, Helen and 300 W, and compare our results with other excellent methods. The parameters are set as follows: $N = 7$, $D = 5$ and $M = 6$.

On the Helen dataset, to keep consistent with other methods, we use 68 landmarks to evaluate our algorithm. From Table 3, we can find that our algorithm achieves a satisfactory result. We believe this all due to the effectiveness of the LPF and the high performance of federated models.

On the 300-W dataset, Table 4 shows the experimental results of our and the competitors' methods. We can see the superiority of our method in the

Table 2. Runtime (in FPS) on 300-W. Our parameter T is set to 2, and the results of other approaches are quoted from the original theses.

Method	CFSS [31]	CFAN [28]	TCDCN [29]	SDM [26]	RCPR [5]	ESR [6]	LBF [21]	Proposed
FPS	25	44	56	70	80	120	320	356

Table 3. Mean errors (percent) on Helen dataset (68 landmarks)

Method	RCPR [5]	SDM [26]	CFAN [28]	CDM [27]	GN-DPM [25]	CFSS [31]	TCDCN [29]	Proposed
Error	5.93	5.50	5.53	9.90	5.69	4.63	4.60	4.98

Table 4. Mean errors (percent) on 300-W dataset (68 landmarks)

Method	Common subset	Challenging subset	Fullset	FPS
CDM [27]	10.10	19.54	11.94	-
DRMF [1]	6.65	19.79	9.22	-
RCPR [5]	6.18	17.26	8.35	80
GN-DPM [25]	5.78	-	-	-
CFAN [28]	5.50	16.78	7.69	44
ESR [6]	5.28	17.00	7.58	120
SDM [26]	5.57	15.40	7.50	70
ERT [14]	-	-	6.40	**1000**
LBF [21]	4.95	11.98	6.32	320
CFSS [31]	4.73	9.98	5.76	25
TCDCN [30]	4.80	8.60	5.54	56
Proposed	**4.61**	**8.38**	**5.35**	33

Fig. 5. Example alignment results on 300-W. The first row shows the challenging sets, the second row shows the Helen dataset of common sets, and the last row shows the LFPW dataset of common sets.

challenging subset as well as the common subset and fullset. As a result of the mechanism of combining various convergent models to predict the final shapes, the proposed method is imperfect relative to other methods in term of speed. Exemplary alignment results of our method are depicted in Fig. 5.

We provide a demo of our method at https://pan.baidu.com/s/1hsqjZqW.

5 Conclusion

In this paper, we propose a novel local probabilistic features based method for face alignment, under cascaded framework. The local probabilistic features contribute to modelling the probability of a test sample reaching each leaf node. This makes the prediction more realistic and theoretical. The cascade framework also obviously enhances the performance of our experiments. By combining various convergent models to overcome the instability of a single one, the regression accuracy of our method is superior to state-of-the-art methods. We demonstrate the efficiency and accuracy of the proposed method on two classical face alignment datasets. Furthermore, it is worth applying the local probabilistic features to many regression problems.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (No. 61472393, No. 61572450 and No. 61303150), the Fundamental Research Funds for the Central Universities (WK2350000002).

References

1. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Robust discriminative response map fitting with constrained local models, vol. 9, no. 4, pp. 3444–3451 (2013)
2. Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 545–552 (2011)
3. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: IEEE International Conference on Computer Vision, pp. 1–8 (2007)

4. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
5. Burgos-Artizzu, X.P., Perona, P., Dollar, P.: Robust face landmark estimation under occlusion. In: IEEE International Conference on Computer Vision, pp. 1513–1520 (2013)
6. Cao, X., Wei, Y., Wen, F., Sun, J.: Face alignment by explicit shape regression. Int. J. Comput. Vis. **107**(2), 177–190 (2014)
7. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Patt. Anal. Mach. Intell. **23**(6), 681–685 (2001)
8. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
9. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models. In: 2006 British Machine Vision Conference, Edinburgh, UK, September, pp. 929–938 (2006)
10. Dantone, M., Gall, J., Fanelli, G., Van Gool, L.: Real-time facial feature detection using conditional regression forests. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2578–2585 (2012)
11. Dollar, P., Welinder, P., Perona, P.: Cascaded pose regression, vol. 238, no. 6, pp. 1078–1085. IEEE (2010)
12. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: a library for large linear classification. J. Mach. Learn. Res. **9**, 1871–1874 (2008)
13. Huang, Z., Zhao, X., Shan, S., Wang, R., Chen, X.: Coupling alignments with recognition for still-to-video face recognition. In: IEEE International Conference on Computer Vision, pp. 3296–3303 (2013)
14. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)
15. Kumar, N., Belhumeur, P., Nayar, S.: FaceTracer: a search engine for large collections of images with faces. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5305, pp. 340–353. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88693-8_25
16. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 679–692. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33712-3_49
17. Lu, C., Tang, X.: Surpassing human-level face verification performance on LRW with Gaussian face. Computer Science (2014)
18. Luo, C., Jiang, C., Yu, J., Wang, Z.: Expressive facial animation from videos. In: IEEE International Conference on Image Processing, pp. 4617–4621 (2014)
19. Messer, K., Matas, J., Kittler, J., Jonsson, K., Luettin, J., Maitre, G.: XM2VTSDB: the extended M2VTS database. In: Proceedings of the Second International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 72–77 (2000)
20. Ramanan, D., Zhu, X.: Face detection, pose estimation, and landmark localization in the wild. In: Computer Vision and Pattern Recognition, pp. 2879–2886 (2012)
21. Ren, S., Cao, X., Wei, Y., Sun, J.: Face alignment at 3000 fps via regressing local binary features. IEEE Trans. Image Process. **25**(3), 1685–1692 (2014)
22. Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: A semi-automatic methodology for facial landmark annotation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 896–903 (2013)
23. Schölkopf, B., Platt, J., Hofmann, T.: Fast discriminative visual codebooks using randomized clustering forests. In: NIPS, pp. 985–992 (2007)

24. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection, vol. 9, no. 4, pp. 3476–3483 (2013)
25. Tzimiropoulos, G., Pantic, M.: Gauss-Newton deformable part models for face alignment in-the-wild. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1851–1858 (2014)
26. Xiong, X., Torre, F.D.L.: Supervised descent method and its applications to face alignment. In: Computer Vision and Pattern Recognition, pp. 532–539 (2013)
27. Yu, X., Huang, J., Zhang, S., Yan, W., Metaxas, D.N.: Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In: IEEE Transactions on Software Engineering, pp. 1944–1951 (2013)
28. Zhang, J., Shan, S., Kan, M., Chen, X.: Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 1–16. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10605-2_1
29. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 94–108. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_7
30. Zhang, Z., Luo, P., Chen, C.L., Tang, X.: Learning deep representation for face alignment with auxiliary attributes. IEEE Trans. Patt. Anal. Mach. Intell. **38**(5), 918 (2016)
31. Zhu, S., Li, C., Chen, C.L., Tang, X.: Face alignment by coarse-to-fine shape searching. In: CVPR, pp. 4998–5006 (2015)